

Maladaptive Behaviours Associated with Generalized Anxiety Disorder: An Item Response Theory Analysis

Alison E.J. Mahoney*

Clinical Research Unit for Anxiety and Depression, University of New South Wales at St Vincent's Hospital, Sydney, NSW, Australia

Megan J. Hobbs

Clinical Research Unit for Anxiety and Depression, University of New South Wales at St Vincent's Hospital, Sydney, NSW, Australia

Jill M. Newby

Department of Psychology, University of New South Wales, Sydney, NSW, Australia

Alishia D. Williams

Department of Clinical and Health Psychology, Utrecht University, The Netherlands

Gavin Andrews

Clinical Research Unit for Anxiety and Depression, University of New South Wales at St Vincent's Hospital, Sydney, NSW, Australia

Background: Cognitive models of generalized anxiety disorder (GAD) suggest that maladaptive behaviours may contribute to the maintenance of the disorder; however, little research has concentrated on identifying and measuring these behaviours. To address this gap, the Worry Behaviors Inventory (WBI) was developed and has been evaluated within a classical test theory (CTT) approach. *Aims:* As CTT is limited in several important respects, this study examined the psychometric properties of the WBI using an Item Response Theory approach. **Method:** A large sample of adults commencing treatment for their symptoms of

* Correspondence to Dr Alison Mahoney, Clinical Research Unit for Anxiety and Depression, University of New South Wales at St Vincent's Hospital, Level 4 O'Brien Centre, 394–404 Victoria Street, Darlinghurst, NSW, Australia, 2010. E-mail: Alison.Mahoney@svha.org.au

GAD ($n = 537$) completed the WBI in addition to measures of GAD and depression symptom severity. **Results:** Patients with a probable diagnosis of GAD typically engaged in four or five maladaptive behaviours most or all of the time in an attempt to prevent, control or avoid worrying about everyday concerns. The two-factor structure of the WBI was confirmed, and the WBI scales demonstrated good reliability across a broad range of the respective scales. Together with previous findings, our results suggested that hypervigilance and checking behaviours, as well as avoidance of saying or doing things that are worrisome, were the most relevant maladaptive behaviours associated with GAD, and discriminated well between adults with low, moderate and high degrees of the respective WBI scales. **Conclusions:** Our results support the importance of maladaptive behaviours to GAD and the utility of the WBI to index these behaviours. Ramifications for the classification, theoretical conceptualization and treatment of GAD are discussed.

Keywords: generalized anxiety disorder, avoidance, psychometric

Introduction

The DSM-5 emphasizes the cognitive and somatic symptoms of generalized anxiety disorder (GAD) (American Psychiatric Association, 2013), which is characterized by excessive and chronic worry and anxiety about everyday events. The worry and anxiety are associated with restlessness, fatigue, difficulty concentrating, irritability, muscle tension and sleep disturbances. Although behavioural symptoms are included in other anxiety diagnostic categories (e.g. social anxiety disorder, panic disorder, and agoraphobia) and were proposed for the DSM-5 GAD classification (Andrews et al., 2010), the relevance of maladaptive behaviours to the disorder is unclear. However, cognitive models of GAD suggest that maladaptive behaviours may contribute to its maintenance and cognitive behavioural treatment protocols direct clinicians to systematically reduce these behaviours during therapy (e.g. Andrews et al., 2016; Dugas et al., 1998; Robichaud, 2013; Wells, 1995, 1999). The development of a valid and reliable index of maladaptive behaviours that are associated with GAD could assist clinicians to monitor treatment progress and researchers to continue to explore important questions about the diagnostic, theoretical and clinical significance of maladaptive behaviours in GAD.

Following the work of Beesdo-Baum and colleagues (2012), Mahoney et al. (2016) developed the Worry Behaviors Inventory (WBI). This 10-item scale indexes maladaptive behaviours associated with symptoms of GAD. Items, such as 'I repeatedly check that things have been done properly', are rated along a 5-point scale based on how frequently patients typically engage in the behaviour (e.g. 0 = *None of the time* to 4 = *All of the time*). Preliminary psychometric evaluations in adults seeking treatment for their symptoms of anxiety and/or depression supported a two-factor structure for the measure (Mahoney et al., 2016, 2018). These factors were labelled *Safety Behaviors* (e.g. checking, watching, planning, reassurance-seeking and controlling others) and *Avoidance* (e.g. avoidance of decision-making, worrying situations, people and activities). Both factors demonstrated evidence of internal consistency, temporal stability, treatment sensitivity, incremental validity and discriminant validity (Mahoney et al., 2018). Supporting the convergent validity of the WBI, the *Safety Behaviors* and *Avoidance* factors were significantly correlated with measures of GAD symptom severity, behavioural inhibition, checking behaviours, and reassurance-seeking (Mahoney et al., 2018). Supporting the divergent validity of the WBI, the *Safety Behaviors* factor was more strongly related to measures of GAD symptom severity than to measures of panic disorder, depression, social

anxiety, health anxiety and personality disorder symptom severity (Mahoney et al., 2016, 2018). However, the WBI subscales demonstrated differential associations with measures of anxiety and depression symptom severity. While the *Safety Behaviors* factor was most strongly related to GAD symptom severity, the *Avoidance* factor was as strongly related to GAD symptom severity as it was to depression and social anxiety symptom severity. This underscores the likely transdiagnostic nature of the behaviours comprising the *Avoidance* scale (Mahoney et al., 2016, 2018).

Existing psychometric analyses of the WBI have been conducted within a classical test theory paradigm. Analysing the psychometric properties of the WBI within an item response theory (IRT) framework could complement existing findings. IRT describes the relationship between the probability of endorsing each item as a function of the item characteristics and the severity of the latent factor that is being measured (e.g. the *Safety Behavior* and *Avoidance* scales of the WBI). This relationship can be graphed using a 'category response function', an example of which is shown in Fig. 1. There are several types of IRT models, with the Samejima IRT model (Samejima, 1969) appropriate for items such as those included in the WBI, that are measured using three (or more) response categories. The x -axis of Fig. 1 indexes the severity of the latent factor being measured and the y -axis indexes the probability of endorsing each category of the respective item. This probability is estimated using two types of parameters: (1) a discrimination (' a ') parameter describes the ability of each item to distinguish between similar degrees of the latent factor being measured; and (2) difficulty (' b ') parameters (i.e. b_1 , b_2 , b_3 and b_4). Each b value indexes the trait level necessary to respond with a 50% probability compared with successive response options (i.e. as the WBI has five response options, *None of the time* to *All of the time*, b_1 is response 0 vs 1, 2, 3, 4; $b_2 = 0, 1$ vs 2, 3, 4; $b_3 = 0, 1, 2$ vs 3, 4; and $b_4 = 0, 1, 2, 3$ vs 4). Ideally, the probability of selecting a lower response option (i.e. *None of the time*) peaks at lower levels of the latent factor and steadily decreases as the latent severity increases, while the probability of selecting a high response option (i.e. *All of the time*) progressively increases as the latent factor increases and then clearly peaks at the higher level of the latent factor (Santor and Ramsay, 1998). Items that do not have a strong ability to discriminate between similar levels of the latent factor will yield a comparable probability of endorsing each of the response options (i.e. the category response functions will be relatively flat across the breadth of the factor).

Using IRT to evaluate the psychometric characteristics of the WBI could provide important information regarding the assessment of maladaptive behaviours that are associated with GAD. First, such analyses would provide detailed information about the relationships between individual WBI items (i.e. specific maladaptive behaviours) and severity, as measured by the *Safety Behaviors* and *Avoidance* latent factors (Reise and Waller, 2009). For instance, are specific behaviours characteristic of mild or severe forms of avoidance? If so, should treatment protocols be tailored to address more severe forms of avoidance earlier in treatment, or should more severe forms of avoidance be considered in classification systems? Second, the relative quality of items, in terms of the amount of psychometric information items provide to the total scale, can be determined (Reise et al., 2005), and this can identify which WBI items best represent the underlying constructs of maladaptive behaviours in GAD. Third, because IRT analyses directly scrutinize the performance of item response categories, recommendations can be made to modify the number of categories that are used to measure each item. This may enhance a measure's clinical utility. The WBI items currently have five response categories ranging from *None of the time* to *All of the time*, but the usefulness of these response categories

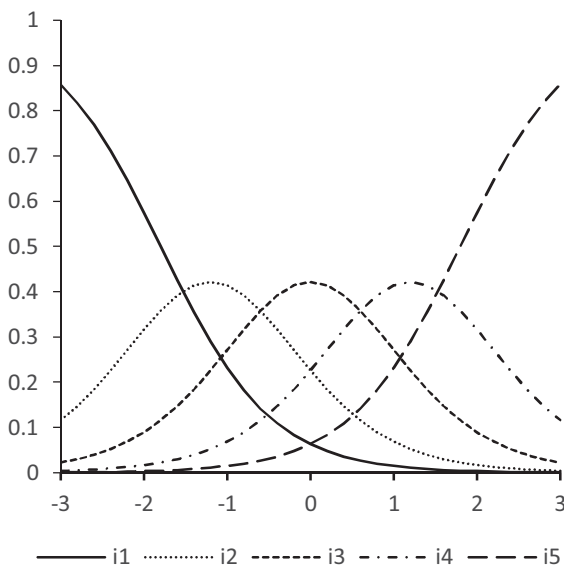


Figure 1. Category response function of a hypothetical ideal WBI item, i , that has five response categories ($i1$, $i2$, $i3$, $i4$, $i5$). This item had a discrimination parameter (a) of 1 and difficulty parameters (b_1 , b_2 , b_3 and b_4) of -1.8 to 1.8 . The vertical axis of the figure indexes the likelihood (ranging from 0 to 1) that a treatment-seeker will report that they experience the respective response option (i.e. ‘none’, ‘a little’, ‘some’, ‘most’, and ‘all of the time’). The horizontal axis indexes the distribution of the latent construct (-3 to 3) measured by the WBI items. This latent psychological construct is typically referred to as theta (θ_j), and is measured on a standard metric. Each of the five trace lines represents one of the five response options that were used to assess the WBI items. The trace line labelled $i1$ (which represents ‘none of the time’) has a greater likelihood of being endorsed by treatment-seekers with lower degrees of latent construct severity, and this likelihood decreases with increasing latent construct severity. That is, if a treatment-seeker has a low degree of latent construct, it is more likely that they will respond to the respective WBI item as ‘none of the time’. The trace lines corresponding to $i2$, $i3$ and $i4$ peak in the mid-ranges of latent construct severity, and the trace line $i5$ (which represents the fifth response option for WBI item i , ‘all the time’), peaks in the high range of latent construct severity. Thus, if a treatment-seeker was experiencing a very high degree of latent construct severity, it would be expected that they would endorse the fifth response category of the hypothetical WBI item i .

has never been examined. Finally, IRT also provides overall test information functions (TIFs) for each factor being measured. Unlike classic test theory, which provides one reliability estimate for a test, IRT can estimate the measurement precision (i.e. reliability) at various degrees of latent severity (e.g. from mild to severe degrees of avoidance). This may be particularly important in treatment settings where a patient’s engagement in maladaptive behaviours is likely to change over the course of time (e.g. Beesdo-Baum et al., 2012). In such cases, users of the WBI need confidence that the measure is reliable across the breadth of the subscales.

Given the diagnostic, theoretical and clinical importance of understanding and measuring maladaptive behaviours associated with GAD, this study conducted an IRT analysis of the WBI in a sample of adults commencing treatment for their symptoms of GAD. We hypothesized that the two-factor structure of the scale would be replicated and that both scales would be significantly related to symptoms of GAD and major depressive disorder (MDD), as has been found in prior evaluations (Mahoney et al., 2016, 2018). However, as previous research has found that the WBI subscales demonstrate differential associations with measures of anxiety and depression, we examined the scales individually and sought to identify the most discriminating items from the *Safety Behaviors* and *Avoidance* subscales, as well as provide indices of the overall functioning of the *Safety Behaviors* and *Avoidance* scales.

Method

Participants

Between 30 August 2013 and 30 October 2015, 537 consecutive patients who were referred for internet-delivered cognitive behaviour therapy (iCBT) for their symptoms of GAD by their general practitioner or mental health professional, completed the WBI as part of their standardized intake assessment (see Andrews and Williams, 2015, for details of the online clinic, ThisWayUpClinic.org.au). This CBT protocol has been validated in two randomized controlled trials (Robinson et al., 2010; Titov et al., 2009), two effectiveness studies (Mewton et al., 2012; Newby et al., 2017) and therapist guides are publicly available (Andrews et al., 2016). Patients were mostly female (63.6%) and in their late thirties (mean (*SD*) = 39.16 (13.83) years, range 18–85 years). Most patients were referred for treatment by their general practitioner (66.5%) or psychologist (16.8%), and resided in urban areas (72.9%) or rural and/or remote communities (24.6%) (patients' rurality was inferred from their postcode and Australian Statistical Geography Standards; Australian Bureau of Statistics, 2013).

Measures

Worry Behaviors Inventory. The WBI is a 10-item, self-report measure that assesses how often respondents typically use maladaptive behaviours to control or prevent worry about everyday concerns. Evidence of internal consistency (total $\alpha = .86$, *Safety Behaviors* subscale $\alpha = .85$, *Avoidance* subscale $\alpha = .75$), temporal stability (total $r = .89$, *Safety Behaviors* subscale $r = .89$, *Avoidance* subscale $r = .74$ over 2–4 weeks) and validity (e.g. convergent/divergent, discriminant, and incremental validity) has been provided (Mahoney et al., 2016, 2018). Internal consistency in the current sample was total $\alpha = .86$, *Safety Behaviors* subscale $\alpha = .85$, *Avoidance* subscale $\alpha = .74$.

Generalized Anxiety Disorder 7-item. The GAD-7 is a 7-item self-report measure of GAD symptom severity experienced in the past 2 weeks (Spitzer et al., 2006). Patients reported how often they experienced symptoms as 'not at all', 'on several days', 'on more than half the days' or 'on nearly every day'. Total scores ≥ 10 indicates a probable GAD diagnosis (sensitivity = 89% and specificity = 82%; Spitzer et al., 2006). Studies support a one-dimensional structure and provide evidence of internal consistency ($\alpha = .92$), temporal stability ($r = .83$), convergent/divergent validity (e.g. correlations with the measures of anxiety, depression), and

criterion validity (e.g. sensitivity/specificity with respect to diagnosis) (Löwe et al., 2008; Spitzer et al., 2006). Internal consistency in the current sample was $\alpha = .89$.

Patient Health Questionnaire-9. The PHQ-9 is a 9-item self-report measure of MDD symptom severity experienced in the past 2 weeks. Patients reported how often they experienced symptoms as ‘not at all’, ‘on several days’, ‘on more than half the days’ or ‘on nearly every day’ with total scores ≥ 10 indicative of a probable diagnosis (Kroenke et al., 2001). Factor analyses support a one- or two-factor structure, and evidence of internal consistency ($\alpha = .86$), temporal stability ($r = .84$ over 48 h), convergent/divergent validity (e.g. correlations with measures of depression and substance use), criterion validity (e.g. sensitivity/specificity with respect to diagnosis) has been reported (Beard et al., 2016; Kroenke et al., 2001, 2010). Internal consistency in the current sample was $\alpha = .88$.

Kessler Psychological Distress Scale. The K-10 is a 10-item measure of psychological distress experienced in the past 2 weeks (Kessler et al., 2002). Patients reported how frequently they had experienced K-10 items as ‘none’, ‘a little’, ‘some’, ‘most’ or ‘all’ of the time, with scores ≥ 20 indicative of clinically significant distress (Kessler et al., 2002). Evaluations support a unidimensional structure, internal consistency ($\alpha = .93$), discriminant validity and treatment sensitivity (Furukawa et al., 2003; Kessler et al., 2002; Mewton et al., 2012; Sunderland et al., 2012). Internal consistency in the current sample was $\alpha = .88$.

Procedure

Before commencing treatment, patients completed an online self-report battery including the WBI, GAD-7, PHQ-9 and K-10. Patients were informed that their data would be collected and their pooled data analysed and published. Patients could opt out of the use of their data via email with no impact on their receipt of treatment. All patients provided electronic informed consent.

Analyses

First, the latent structure of the WBI items was assessed in order to confirm whether unidimensional or multidimensional IRT analyses were indicated (Reckase, 2009). This structural analysis was also used to estimate the correlations between the WBI factor scores and the summed scores of the GAD-7, PHQ-9 and K-10, thereby providing further data regarding the convergent and divergent validity of the WBI. Second, the IRT properties of the WBI items were evaluated. Finally, the test characteristics and reliability of the *Safety Behaviors* and *Avoidance* scales were estimated.

Latent structure of the WBI

Based on the existing psychometric literature, two confirmatory factor analyses (CFAs) were specified: Model 1 used a single factor to explain the co-occurrence of the WBI items, while Model 2 used two factors to explain the co-occurrence of the WBI items (these factors reflected the previously identified *Safety Behaviors* and *Avoidance* scales). CFA models were estimated using a weighted least squares mean and variance estimator that is suitable for categorical/ordinal variables in the MPlus v. 8 software package (Muthén and Muthén, 1998).

The residual correlation matrix produced by the single-factor CFA was investigated to evaluate the local independence of the WBI item bank. The criterion for the violation of the local independence was defined as a residual correlation greater than 0.2 with any of the remaining test items (Reeve et al., 2007). The fit of the CFA models was assessed with reference to the comparative fit index (CFI), Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA). Hu and Bentler (1998) suggest that CFI and TLI $\geq .95$ and RMSEA $< .05$ indicate that the estimated model provides excellent fit to the data. MacCallum et al. (1996) advise that RMSEA values in the range of .08 to .10 indicate mediocre fit. Item fit to the selected IRT model was estimated via the $S-X^2$ statistic using the FlexMIRT software program (Cai, 2017). Possible item misfit was identified when the p -value associated with the $S-X^2$ statistic was $< .05$ (Kang and Chen, 2011; Orlando and Thissen, 2000).

IRT parameters of the WBI items

Structural analyses of the WBI indicated that unidimensional rather than multidimensional IRT methods were appropriate (detailed further below in the Results section), and so separate unidimensional models were used to estimate the IRT parameters of each WBI scale. The Samejima (1969) IRT graded response model was therefore used to evaluate the relationship between patients' WBI severity and the likelihood that they would endorse each item. Discrimination and difficulty parameters were calculated using the recommendations of Muthén (2001) (e.g. $a = \text{loading}/\sqrt{(1 - \text{loading}^2)}$, $b = \text{threshold}/\text{loading}$). Category response functions (CRFs) were calculated using standard formula (Reckase, 2009; Samejima, 1969), and then graphed in Excel 2013.

Test information functions (TIF) and the reliability of the Safety Behavior and Avoidance Scales

In IRT, the amount of test information differs as a function of latent severity (θ) and the a - and b -parameters of the test items. Item information functions (IIFs) were calculated using standardized formulae and summated to produce the TIF of the respective scale. The standard error of the estimate (SEE) is an inverse function of this TIF (i.e. $SEE = 1/\sqrt{I(\theta)}$). The larger the information values and lower the SEE the more precise the measurement of the WBI. To examine the precision of the WBI scales at clinically meaningful levels, we converted the SEE into the reliability co-efficient used in classic psychometric evaluations for different degrees of latent severity (e.g. $r = 1 - SEE^2$) (Thissen and Wainer, 2001).

Results

Sample characteristics

Measures of symptom severity indicated that patients were characterized by substantial rates of probable disorder. Sixty-one per cent of patients met criteria for probable GAD [GAD-7 mean (SD) = 11.42 (5.17); median (IQR) = 11.00 (9)], 45% for probable MDD [PHQ-9 mean (SD) = 9.50 (5.75); median (IQR) = 9.00 (8)], and 78% reported clinically significant levels of psychological distress [K-10 mean (SD) = 25.70 (6.70); median (IQR) = 26.00 (11)].

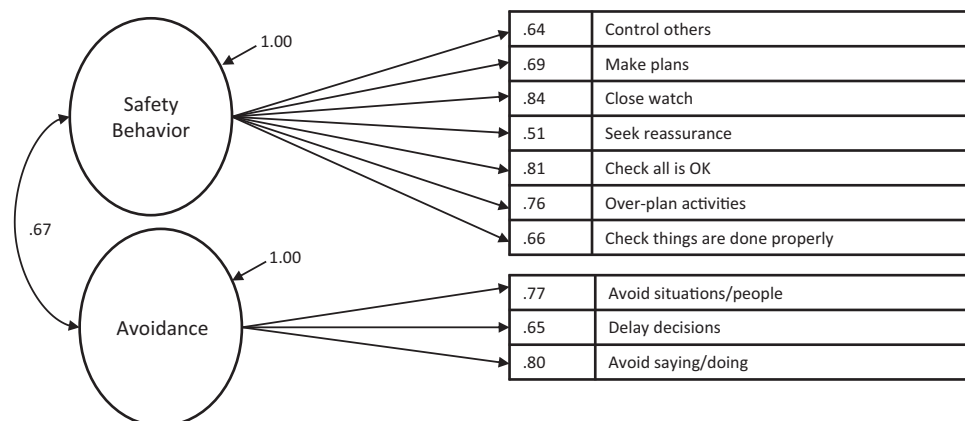


Figure 2. Two-factor model derived from the confirmatory factor analysis of the WBI, indicating standardized factor loadings for each item and inter-construct correlations; ‘Safety behavior’ = the latent factor underpinning the WBI items that comprise the *Safety Behaviors* scale; ‘Avoidance’ = the latent factor underpinning the WBI items that comprise the *Avoidance* scale.

Endorsement of the WBI items

Endorsement of the WBI items is shown in Table 1. Hypervigilance (item 3), checking (item 6) and behavioural avoidance (items 4 and 8) were the most frequently endorsed items (i.e. those most frequently endorsed at *Most* or *All of the time*). Seventy-eight per cent of the sample endorsed at least one of the ten WBI items as *Most* or *All of the time*. Of those with probable GAD, 88.1% reported engaging in at least one maladaptive behaviour most or all of the time, with an average experiencing four or more WBI behaviours most or all of the time [mean (*SD*) = 4.10 (2.79), median (IQR) = 4 (4), mode = 5, range = 0–10].

Latent structure of the WBI and relationships with the GAD-7, PHQ-9 and K-10

Model 1 used a single factor to explain the co-occurrence of the WBI items. Investigation of the residual correlation matrix indicated that no WBI items demonstrated problematic local dependence (i.e. no correlations > 0.2). Although Model 1 did not demonstrate ideal fit for the data (CFI = .90, TLI = .87, RMSEA = .16), it represented the most parsimonious model and was used as the baseline standard against which to compare Model 2. Model 2 demonstrated better fit to the data than Model 1 (CFI = .96, TLI = .95, RMSEA = .09). This model is shown in Fig. 2. The standardized factor loadings were all significant at the $p < .001$ level, and for the *Safety Behaviors* and *Avoidant* factors, ranged from .51 to .84, suggesting that all WBI items loaded on their respective factor within a moderate to strong range. When item fit was investigated, the p -values associated with the $S-X^2$ statistic for each item in the *Safety Behaviors* and *Avoidant* scales were > .05, suggesting that misfit was not likely. Within our sample commencing treatment for excessive worry, the factor scores of the *Safety Behaviors* and *Avoidance* scales were significantly related to the summed scores of the GAD-7 (*Safety Behaviors* $r = .50$, *Avoidance* $r = .48$, p -values < .001), PHQ-9 (*Safety Behaviors* $r = .48$,

Table 1. Endorsement and IRT parameters of the WBI items among patients prescribed internet CBT for GAD in routine clinical care

	Endorsement of WBI items (%)					IRT parameters				
	None	A little	Some	Most	All	<i>a</i>	<i>b</i> ₁	<i>b</i> ₂	<i>b</i> ₃	<i>b</i> ₄
Safety behaviours										
1. I try to control what other people do or think (e.g. stop others doing things, tell others to be careful, give advice, over-protect others, do things for others)	18.4	25.1	26.6	21.2	8.6	0.8	-1.4	-0.3	0.8	2.1
2. I make plans 'just in case'	19.4	27.9	22.7	22.3	7.6	1.1	-1.2	-0.1	0.7	2.0
3. I keep a close watch for anything bad that could happen	12.5	18.1	28.1	29.2	12.1	1.5	-1.4	-0.6	0.3	1.4
5. I seek reassurance from sources of information (e.g. personal records, internet, reviews, books)	18.8	25.3	29.3	19.9	6.7	0.6	-1.8	-0.3	1.3	3.0
6. I check to make sure nothing bad has happened or that everything is OK	12.8	24.4	23.8	31.1	7.8	1.3	-1.4	-0.4	0.4	1.8
9. I over-plan activities (e.g. prepare for all possible bad outcomes, have a plan B, plan every step of an activity)	21.4	29.2	22.9	18.2	8.2	1.2	-1.0	0.0	0.8	1.8
10. I repeatedly check that things have been done properly	11.7	27.9	29.2	23.3	7.8	0.9	-1.8	-0.4	0.8	2.2
Avoidance behaviours										
4. I avoid situations or people that worry me	8.2	18.4	33.0	30.2	10.2	1.0	-2.0	-0.9	0.4	1.8
7. I delay making decisions about things or get other people to make decisions for me	16.0	26.6	31.5	19.9	6.0	0.8	-1.6	-0.3	1.1	2.5
8. I avoid saying or doing things that worry me	8.4	22.7	31.1	31.8	6.0	1.9	-1.6	-0.6	0.4	1.8

Items are listed in order of administration, but segregated by subscale. IRT, item response theory; GAD, generalized anxiety disorder; CBT, cognitive behaviour therapy; WBI, Worry Behaviours Inventory; None, none of the time; A little, a little of the time; Some, some of the time; Most, most of the time; All, all of the time; *a*, discrimination parameter (higher values indicate a greater capacity to discriminate between similar levels of the latent factor); *b*₁, *b*₂, *b*₃ and *b*₄, difficulty parameters (*b*₁ indexes the point on the latent factor at which the probability of endorsing the first response option (e.g. *None of the time*) is 50% or greater, *b*₂ indexes this for the *A little of the time* response option, and so forth for the *Some*, *Most* and *All of the time* response options. Ideally, these difficulty parameters spread across the range of the latent trait, that is from -3 to +3 of θ).

Avoidance $r = .49$; p -values $< .001$), and K-10 (*Safety Behaviors* $r = .47$, *Avoidance* $r = .51$, p -values $< .001$). We also observed that the GAD-7 and PHQ-9 total scores were significantly related to each other ($r = .67$, $p < .001$).

IRT characteristics of the WBI items

Table 1 provides the IRT parameters of the 10 WBI items and the respective category response functions are given in Fig. 3 (i.e. Fig. 3 is a graphical representation of the IRT parameters presented in Table 1). The most discriminating *Safety Behaviors* items were 'I keep a close watch for anything bad that could happen' ($a = 1.48$) and 'I check to make sure nothing bad has happened or that everything is OK' ($a = 1.32$), whereas the most discriminating *Avoidance* item was 'I avoid saying or doing things that worry me' ($a = 1.90$). Compared with the other items, these WBI items were the best able to discriminate between adults with similar degrees of safety-seeking and avoidant behaviours.

Given the polytomous nature of the WBI items, we expected that the difficulty parameters would be distributed across the breadth of the *Safety Behaviors* and *Avoidance* factors (that is across -3 to $+3$ of the factors). This was not the case for all items. For instance, consider the category response function in Fig. 3 for item 3 ('Close watch'); four of its five categories (specifically, every response except the *A little of the time* category) were more likely than all of the remaining categories to be endorsed over a unique portion of the *Safety Behavior* factor (that is, one can observe that four of the five curves peak above the remaining curves). Similarly, consider the category response function in Fig. 3 for item 8 ('Avoid saying/doing'); here each of the five response categories (*None of the time* through to *All of the time*) peaked over a unique portion of the *Avoidance* scale (again, the peak of each category response curve is above all the remaining curves). The capacity of each category that was used to assess the remaining WBI items was sub-optimal, with the multiple response categories failing to peak over a unique portion of the respective latent factors. For example, the category response function of 'Seek reassurance' (item 5) in Fig. 3 shows that the probability of endorsing categories other than the *None* or *All of the time* responses provided little psychometric value. This is because the likelihood of endorsing the other categories (*A little*, *Some* and *Most of the time*) did not exceed the likelihood of endorsing the latter response categories (one can observe that the curves for the *A little*, *Some* and *Most of the time* responses did not peak above the curves for the *None* or *All of the time* responses).

Test Information Functions and Reliability of the Safety and Avoidance Behavior Scales

Figure S1 in the online supplementary material provides the IIFs for the WBI items. As expected, the IIFs for most items indicated that measurement error was greater than the information provided by the item (this was not the case for items 3, 6 and 8), and items are thus compiled to form scales. The TIFs for the *Safety Behavior* and *Avoidance* scales are shown in Fig. 4 and are the cumulative function of the respective IIFs (i.e. the TIF for the *Safety Behavior* is the cumulative function of the IIFs for the seven WBI items that form the *Safety Behavior* scale). The *Safety Behavior* and *Avoidance* scales yielded the most information (e.g. the highest precision of measurement) between one standard deviation (*SD*) above and below the mean. Importantly, both scales provide more precision than error across the full range of the factor. Indeed, between $2SD$ below the mean and $2.5SD$ above the mean the *Safety*

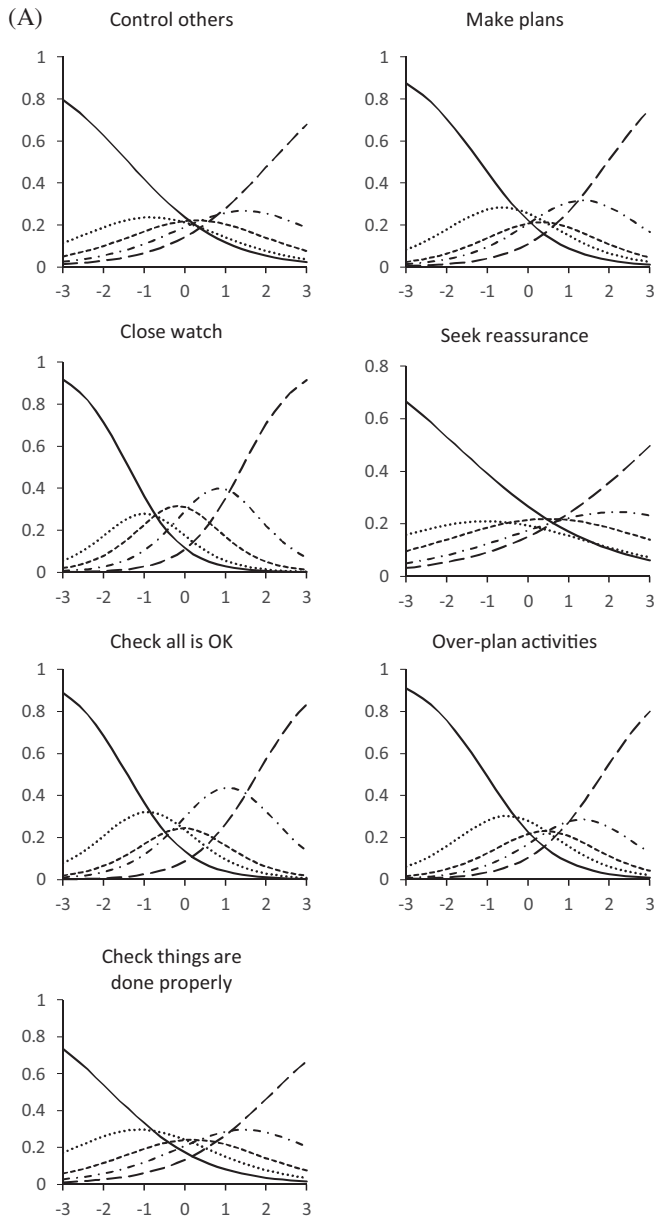


Figure 3. (A) Matrix plot of the category response functions for the *Safety Behaviors* scale. (B) Matrix plot of the category response functions the *Avoidance* scale. For both panels, the y-axis measures the probability of endorsing each response category and the x-axis measures the severity of the latent factor being measured on a standard normal metric. The continuous line, small dashed line, medium dashed line, broken line and the long dashed line plot the relationship between the probability of endorsing the ‘None of the time’, ‘A little of the time’, ‘Some of the time’, ‘Most of the time’ and ‘All of the time’ categories of the WBI items, respectively.

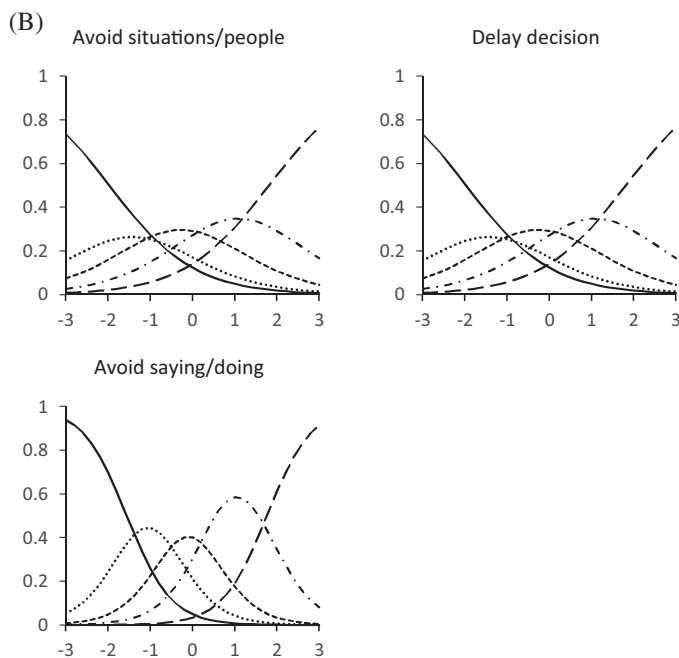


Figure 3. (Continued)

Behaviors scale had good reliability ($R = 0.75$ – 0.75); and between 1 and $0.5SD$ below the mean to $1SD$ above the mean the *Avoidance* scale demonstrated acceptable reliability ($R = 0.70$ – 0.70) (detailed reliability estimates for each level of θ in .2 increments are available in Table S2 of the online supplementary material).

Discussion

Maladaptive behaviours, such as checking and avoiding worry-provoking situations, have been theorized to contribute to the maintenance of GAD, and as such, are targeted in cognitive behaviour therapy for GAD (Andrews et al., 2016; Dugas et al., 1998; Robichaud, 2013; Wells, 1995, 1999). The introduction of maladaptive behaviours was considered for the DSM-5 classification of GAD, but little empirical data were available at the time to determine which maladaptive behaviours were most relevant to GAD (Andrews et al., 2010). The development of the Worry Behaviors Inventory (WBI) sought to advance our understanding of these behaviours by providing a brief self-report index of clinically meaningful maladaptive behaviours associated with GAD. The preliminary psychometric evaluations of the WBI supported the reliability and validity of the scale, but were exclusively conducted within the context of classical test theory. The current study undertook an IRT analysis of the WBI and provides important additional information about the quality and reliability of the WBI items.

As predicted, exploratory and confirmatory factor analyses of the WBI supported the two-factor structure found in previous studies (Mahoney et al., 2016, 2018). Indeed, the two-factor

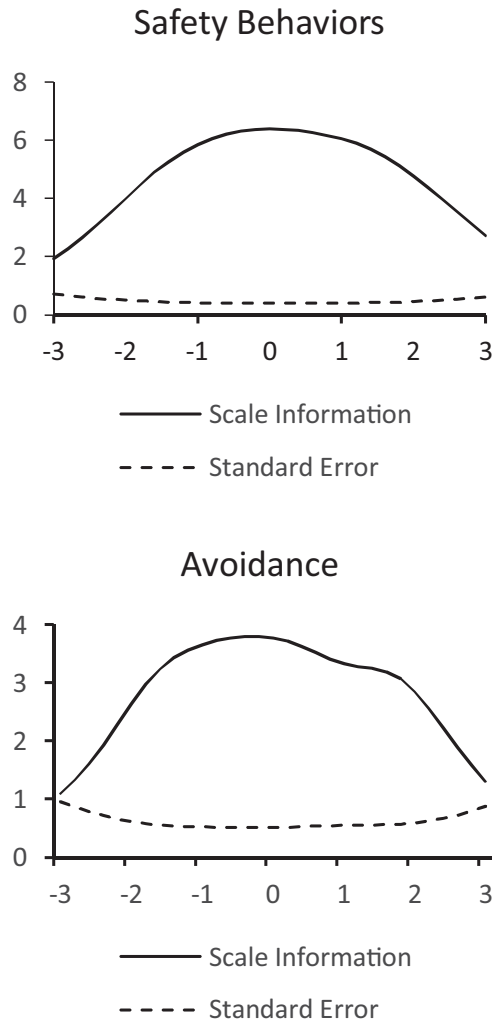


Figure 4. Matrix plot of the test information functions (TIFs) for the two subscales of the WBI. For each function, the horizontal axis indicates the distribution of the latent construct (θ_j) in scale scores (ranging from -3 to 3), and the vertical axis indicates the test information available.

structure provided a closer fit to the current data than previous data. This is probably the result of sample variations because previous samples have consisted of more heterogeneous clinical samples seeking treatment for their symptoms of anxiety and/or depression (Mahoney et al., 2016, 2018). Extending previous psychometric evaluations regarding the convergent and divergent validity of the WBI scales, the *Safety Behavior* and *Avoidance* latent factors were found to significantly correlate with psychological distress and symptoms of GAD and MDD. Both behavioural constructs appeared to be as strongly related to GAD as to MDD symptom severity in our sample undertaking treatment for GAD. Although these findings

support a transdiagnostic conceptualization of these maladaptive behaviours, it is yet to be seen if the strength of these associations would be different in samples who were primarily seeking treatment for depression. Such evaluations would provide a more rigorous examination of the possible transdiagnostic nature of these maladaptive behaviours. Our findings are also consistent with previous studies demonstrating the considerable overlap between the latent traits that underpin GAD and MDD symptoms (Zbozinek et al., 2012). Our data suggest that GAD symptoms represent a related but distinct construct to that of MDD symptoms, and both GAD and MDD (which predominantly consist of cognitive, emotional and somatic symptoms) are significantly associated with maladaptive behaviours.

The majority of the current sample (78%) reported that they typically engaged in at least one maladaptive behaviour most or all of the time when responding to their worries about everyday events. This proportion was higher in patients with probable GAD where almost 90% of patients reported that they frequently engaged in at least one maladaptive behaviour to manage their worry. These findings support the relevance of maladaptive behaviours in adults seeking treatment for their symptoms of GAD and are consistent with cognitive theories of GAD (Andrews et al., 2016; Dugas et al., 1998; Wells, 1995, 1999). A subset of the WBI items were highly endorsed by the current sample: 'I keep a close watch for anything bad that could happen' (item 3); 'I check to make sure nothing bad has happened or that everything is OK' (item 6); 'I avoid situations or people that worry me' (item 4) and 'I avoid saying or doing things that worry me' (item 8). Interestingly, hypervigilance comprised part of the DSM-III classification of GAD and pathological checking behaviours have, to date, been the most studied maladaptive behaviour associated with GAD (American Psychiatric Association, 1987; Coleman et al., 2011; Schut et al., 2001; Tallis and de Silva, 1992; Townsend et al., 1999). Importantly, these items performed well during IRT analyses.

IRT findings suggested that the most discriminating item along the full range of the *Safety Behaviors* latent construct was item 3 (Keep a close watch), whereas the most discriminating item along the full range of the *Avoidance* latent construct was item 8 (Avoid saying or doing things). Compared with other items, these items provided the most psychometric information about their respective latent factors. The capacity of the remaining WBI item categories to measure information across the full range of the *Safety Behaviors* and *Avoidance* constructs was sub-optimal, with multiple response categories in most items failing to provide unique psychometric information about their respective latent factor. We anticipated that the difficulty parameters of the response categories would be spread across the latent factors because of the polytomous response format of the WBI items, but this was not the case. Rather, the category response functions for most items suggested that a more appropriate response option would be dichotomous (i.e. the patient generally does or does not engage in the particular maladaptive behaviour in response to worry). From a clinical standpoint, modification of the response options may result in a simplified and more rapid administration, but it is unclear if such a change would negatively impact on the treatment sensitivity of the measure or the clinical richness of responses which may be used for case conceptualization and treatment planning.

Nevertheless, the analysis of overall test information estimates suggested that the *Safety Behaviors* and *Avoidance* scales demonstrated more precision of measurement than error across the full range of the scales. Extending previous findings regarding the internal consistency of the WBI scale (Mahoney et al., 2016, 2018), this study found that the *Safety Behaviors* scale provided reliable measurement across a broad range of the scale ($R = 0.75\text{--}0.79$ at $\pm 2SD$ of θ). The *Avoidance* scale provided adequate reliability of measurement around the

mean range of the scale ($R = 0.72\text{--}0.70$ at $\pm 1.2SD$ of θ). As such, researchers and clinicians can be assured of the precision of measurement for most patients as they progress over time and/or treatment.

There are a number of noteworthy clinical and theoretical ramifications of this study. Together with previous psychometric analyses, there is convergence about which WBI items perform most strongly. Hypervigilance and checking behaviours, as well as avoidance of saying or doing things that are worrisome, appear to be the most relevant maladaptive behaviours associated with GAD, and the best able to discriminate between adults with low, moderate and high degrees of *Safety Behaviors* and *Avoidance* (Mahoney et al., 2016). We found that patients with a probable diagnosis of GAD reported that they typically engaged in four or five maladaptive behaviours most or all of the time in an attempt to prevent, control or avoid worrying about everyday concerns. These behaviours need to be identified and treated if they contribute to the maintenance of the disorder as theorized (Dugas et al., 1998; Wells, 1999). Current DSM-5 diagnostic criteria for many of the anxiety disorders specifies that a significant change in behaviour occurs in relation to the anxiety, whether it be avoidance of social and performance situations in social anxiety disorder or avoidance of exercise as specified in the panic disorder diagnostic category for instance (American Psychiatric Association, 2013). The current findings are consistent with the proposals put forth for the DSM-5 classification of GAD (Andrews et al., 2010), and are supportive of the further consideration of behavioural criteria in the GAD diagnostic category. However, it is yet to be seen whether the addition of behavioural features to the diagnostic criteria of GAD would significantly enhance the reliability, validity and clinical utility of the classification.

Our findings need to be interpreted within the limitations of this study. We did not administer structured diagnostic interviews. As such, we were unable to confirm the diagnostic profile of the sample and conclusions regarding the clinical and theoretical ramifications of this study must be moderated accordingly. However, patients in this study were referred by their clinician and elected to complete treatment for their symptoms of GAD, and this maximizes the ecological validity of the study. Furthermore, the inclusion of sub-threshold cases enhances IRT analyses where, ideally, a broad range of the latent trait severity should be sampled.

Conclusion

Most adults seeking internet-delivered cognitive behaviour therapy for GAD report that they characteristically engage in at least one maladaptive behaviour most or all of the time to manage their worry about everyday concerns. However, when GAD symptoms are clinically severe, patients are most likely to frequently engage in four or five maladaptive behaviours. The WBI represents one reliable and valid self-report scale to assess these clinically meaningful behaviours, and current findings suggest that a dichotomous response format to items (i.e. the patient generally does or does not engage in the unhelpful behaviour) may be most suitable.

Acknowledgements

None.

Ethical statement: All authors have abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the APA. Data from the clinical sample were collected within the

Quality Assurance Activities of St Vincent's Hospital, Sydney Australia. All patients were informed and provided electronic informed consent that their data would be collected, their pooled data analysed and published in scientific journals. Patients could opt out of the use of their data for these purposes via email with no impact on their receipt of treatment.

Conflicts of interest: The authors have no conflicts of interest with respect to this publication.

Financial support: This research received no specific grant from any funding agency, commercial or not-for-profit sectors. A.M. is supported by an Australian Government Research Training Program Scholarship. J.N. is supported by a NHMRC Early Career Research Fellowship (1033787).

Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1352465818000127>

References

- Andrews, G., Hobbs, M. J., Borkovec, T. D., Beesdo, K., Craske, M. G., Heimberg, R. G. et al.** (2010). Generalised worry disorder: a review of DSM-IV generalised anxiety disorder and options for DSM-V. *Depression and Anxiety*, *27*, 134–147.
- Andrews, G., Mahoney, A. E. J., Hobbs, M. J. and Genderson, M.** (2016). *Treatment of Generalized Anxiety Disorder: Therapist Guides and Patient Manual*. Oxford: Oxford University Press.
- Andrews, G. and Williams, A. D.** (2015). Up-scaling internet-based cognitive behavioral therapy (iCBT) for depression: a model for dissemination into primary care. *Clinical Psychology Review*, *41*, 40–48.
- American Psychiatric Association** (1987). *Diagnostic and Statistical Manual for Mental Disorders* (3rd edn). Washington, DC.
- American Psychiatric Association** (2013). *Diagnostic and Statistical Manual for Mental Disorders* (5th edn). Washington, DC.
- Australian Bureau of Statistics** (2013). *Australian Statistical Geography Standard (ASGS)* (vol. 4). Canberra, Australia: Australian Bureau of Statistics.
- Beard, C., Hsu, K. J., Rifkin, L. S., Busch, A. B. and Björgvinsson, T.** (2016). Validation of the PHQ-9 in a psychiatric sample. *Journal of Affective Disorders*, *193*, 267–273.
- Beesdo-Baum, K., Jenjahn, E., Hofler, M., Lueken, U., Becker, E. S. and Hoyer, J.** (2012). Avoidance, safety behavior, and reassurance-seeking in generalized anxiety disorder. *Depression and Anxiety*, *29*, 948–957.
- Cai, L.** (2017). *flexMIRTR* version 3.51: Flexible multilevel multidimensional item analysis and test scoring [computer software]. Chapel Hill, NC: Vector Psychometric Group.
- Coleman, S. L., Pieterfesa, A. S., Holaway, R. M., Coles, M. E. and Heimberg, R. G.** (2011). Content and correlates of checking related to symptoms of obsessive compulsive disorder and generalised anxiety disorder. *Journal of Anxiety Disorders*, *25*, 293–301.
- Dugas, M. J., Gagnon, F., Ladouceur, R. and Freeston, M. H.** (1998). Generalized anxiety disorder: a preliminary test of a conceptual model. *Behavior Research and Therapy*, *36*, 215–226.
- Furukawa, T. A., Kessler, R. C., Slade, T. and Andrews, G.** (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine*, *33*, 357–362.

- Hu, L. T. and Bentler, P. M.** (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453.
- Kang, T. and Chen, T. T.** (2011). Performance of the generalized S-X2 item fit index for the graded response model. *Asia Pacific Educational Review*, 12, 89–96.
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. T. et al.** (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32, 959–976.
- Kroenke, K., Spitzer, R. and Williams, J. B.** (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606–613.
- Kroenke, K., Spitzer, R., Williams, J. B. and Löwe, B.** (2010). The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General Hospital Psychiatry*, 32, 345–359.
- Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W. et al.** (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, 46, 266–274.
- MacCallum, R. C., Browne, M. W. and Sugawara, H. M.** (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.
- Mahoney, A., Hobbs, M. J., Newby, J. M., Williams, A. D., Sunderland, M. and Andrews, G.** (2016). The Worry Behaviors Inventory: assessing the behavioral avoidance associated with generalized anxiety disorder. *Journal of Affective Disorders*, 203, 256–264.
- Mahoney, A., Hobbs, M. J., Newby, J. M., Williams, A. D. and Andrews, G.** (2018). Psychometric properties of the Worry Behaviors Inventory: replication and extension in a large clinical and community sample. *Behavioral and Cognitive Psychotherapy*, 46, 84–100.
- Mewton, L., Wong, N. and Andrews, G.** (2012). The effectiveness of internet cognitive behavioral therapy for generalized anxiety disorder in clinical practice. *Depression and Anxiety*, 29, 843–849.
- Muthén, B.** (2001). Statmodel discussion: comment 16 July 2001, 1035 regarding the converting MPlus output for polytomous items to item response theory parameters of Samejima's graded response model. Available at: <http://www.statmodel.com/discussion/messages/23/35.html> (accessed 21 August 2017).
- Muthén, L. K. and Muthén, B. O.** (1998– 2015). *Mplus User's Guide*, 7th edn. Los Angeles, CA: Muthén and Muthén.
- Newby, J. M., Mewton, L. and Andrews, G.** (2017). Transdiagnostic versus disorder-specific internet-delivered cognitive behavior therapy for anxiety and depression in primary care. *Journal of Anxiety Disorders*, 46, 25–34.
- Orlando, M. and Thissen, D.** (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50–64.
- Reckase, M. D.** (2009). *Multidimensional Item Response Theory*. New York, NY: Springer.
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A. et al.** (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*, 45 (suppl 1), S22–31.
- Reise, S. P., Ainsworth, A. T. and Haviland, M. G.** (2005). Item response theory: fundamentals, application, and promise in psychological research. *Current Directions in Psychological Science*, 14, 95–101.
- Reise, S. P. and Waller, N. G.** (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 5, 27–48.
- Robichaud, M.** (2013). Cognitive behavior therapy targeting intolerance of uncertainty: application to a clinical case of generalized anxiety disorder. *Cognitive and Behavioral Practice*, 20, 251–263.

- Robinson, E., Titov, N., Andrews, G., McIntyre, K., Schwencke, G. and Solley, K.** (2010). Internet treatment for generalised anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. *PLoS ONE*, *5*, e10942.
- Samejima, F.** (1969). *Estimation of latent ability using a response pattern of graded scores*. Psychometric Monograph No. 17. Psychometric Society: Richmond, VA.
- Santor, D. A. and Ramsay, J. O.** (1998). Progress in the technology of measurement: Application of item response theory. *Psychological Assessment*, *10*, 345–359.
- Schut, A. J., Castonguay, L. G. and Borkovec, T. D.** (2001). Compulsive checking behaviors in generalised anxiety disorder. *Journal of Clinical Psychology*, *57*, 705–715.
- Spitzer, R. L., Kroenke, K., Williams, J. B. and Löwe, B.** (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097.
- Sunderland, M., Mahoney, A. and Andrews, G.** (2012). Investigating the factor structure of the Kessler 10 Psychological Distress Scale in community and clinical samples of the Australian population. *Journal of Psychopathology and Behavioral Assessment*, *34*, 253–259.
- Tallis, F. and de Silva, P.** (1992). Worry and obsessional symptoms: a correlational analysis. *Behavior Research and Therapy*, *30*, 103–105.
- Thissen, D. and Wainer, H.** (eds) (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Titov, N., Andrews, G., Robinson, E., Schwencke, E., Johnston, L., Solley, K. and Choi, I.** (2009). Clinician-assisted Internet-based treatment is effective for generalized anxiety disorder: a randomized controlled trial. *Australian and New Zealand Journal of Psychiatry*, *43*, 905–912.
- Townsend, M. H., Weissbecker, K. A., Barbee, J. G., Peña, J. M., Snider, L. M., Tynes, L. L. et al.** (1999). Compulsive behaviors in generalised anxiety disorder. *Journal of Nervous and Mental Disease*, *187*, 697–699.
- Wells, A.** (1995). Meta-cognition and worry: a cognitive model of generalized anxiety disorder. *Behavioral and Cognitive Psychotherapy*, *23*, 301–320.
- Wells, A.** (1999). A metacognitive model and therapy for generalized anxiety disorder. *Clinical Psychology and Psychotherapy*, *6*, 86–95.
- Zbozinek, T. D., Rose, R. D., Wolitzky-Taylor, K. B., Sherbourne, C., Sullivan, G., Stein, M. B. et al.** (2012). Diagnostic overlap of generalized anxiety disorder and major depressive disorder in a primary care sample. *Depression and Anxiety*, *29*, 1065–1071.